Review Article

Drug- and Herb-Induced Liver Injury in Clinical and Translational Hepatology: Causality Assessment Methods, Quo Vadis?

Rolf Teschke[1], Axel Eickhoff[1] and Johannes Schulze[2]

[1]Department of Internal Medicine II, Division of Gastroenterology and Hepatology, Klinikum Hanau, Academic Teaching Hospital of the Medical Faculty, Goethe University Frankfurt/ Main, Germany; [2]Institute of Industrial, Environmental and Social Medicine, Medical Faculty, Goethe University Frankfurt/Main, Germany

Abstract

Drug-induced liver injury (DILI) and herb-induced liver injury (HILI) are typical diseases of clinical and translational hepatology. Their diagnosis is complex and requires an experienced clinician to translate basic science into clinical judgment and identify a valid causality algorithm. To prospectively assess causality starting on the day DILI or HILI is suspected, the best approach for physicians is to use the Council for International Organizations of Medical Sciences (CIOMS) scale in its original or preferably its updated version. The CIOMS scale is validated, liver-specific, structured, and quantitative, providing final causality grades based on scores of specific items for individual patients. These items include latency period, decline in liver values after treatment cessation, risk factors, co-medication, alternative diagnoses, hepatotoxicity track record of the suspected product, and unintentional re-exposure. Provided causality is established as probable or highly probable, data of the CIOMS scale with all individual items, a short clinical report, and complete raw data should be transmitted to the regulatory agencies, manufacturers, expert panels, and possibly to the scientific community for further refinement of the causality evaluation in a setting of retrospective expert opinion. Good-quality case data combined with thorough CIOMS-based assessment as a standardized approach should avert subsequent necessity for other complex causality assessment methods that may have inter-rater problems because of poor-quality data. In the future, the CIOMS scale will continue to be the preferred tool to assess causality of DILI and HILI cases and should be used consistently, both prospectively by physicians, and retrospectively for subsequent expert opinion if needed. For comparability and international harmonization, all parties assessing causality in DILI and HILI cases should attempt this standardized approach using the updated CIOMS scale.

Drugs and herbs are commonly used to cure, stabilize, and prevent disease, or to retain or improve general health conditions. However, drug/herb treatments may be associated with adverse drug reactions (ADR)[1] or adverse herb reactions (AHR).[2] Although most publications provide sufficient evidence that the assumed products likely caused the reactions observed in various organs, this does not necessarily apply to liver ADRs and AHRs. Both challenges and pitfalls in causality attribution have emerged during case assessments for drug-induced liver injury (DILI)[3,4] and herb-induced liver injury (HILI),[5,6] as the clinical signs are similar in both conditions.[7]

DILI and HILI are typical diseases of clinical and translational hepatology in a broader sense, as the complex process of their diagnosis requires experience when translating basic science into clinical judgment, including causality evaluation.[3–7] The physician's results may then be reported and lead to regulatory actions, provided causality has been established. The overall translational process ends with basic science conclusions and pharmacovigilance decisions to prevent future damage. Therefore, key requirements for DILI and HILI are valid evaluations of suspected cases, applying appropriate causality assessment algorithms.

In this review, we address issues of liver-specific causality assessment methods (CAMs) in DILI and HILI cases and present considerations for future strategies.

Types of causality assessment methods

There is considerable interest in both liver-specific and liver-unspecific CAMs,[1–40] to be applied prospectively or retrospectively.[7,28] Methods classified as prospective may be used on the day that DILI or HILI diagnosis is suspected and thereafter, providing a strategy for physicians to gather all required items while the disease is ongoing. A prospective approach is the only possible tool for physicians treating patients with suspected DILI or HILI to carry out timely assessment of causality. By contrast, retrospective assessment methods commonly require an expert team providing evaluation delayed by months or years. Thus, results are not present at therapy for the treating physician, and it is not possible to collect additional data.

Liver specificity

Liver-specific CAMs may be used primarily for prospective or retrospective evaluations (Table 1).[7–19,28] The first pragmatic CAM designed specifically for liver injury cases was published in 1988[8] and formed a sophisticated basis for subsequent algorithms.[9–16] This early CAM was the result of consensus meetings organized by Roussel Uclaf, and initially had no name.[8] For reasons of clarity and transparency, this method was later referred to as the qualitative RUCAM (Roussel Uclaf Causality Assessment Method)[9] and considered to have qualitative rather than quantitative criteria.[8,9] In 1990, progress was made on a standard definition of DILI under the auspices of the Council for International Organizations of Medical Sciences (CIOMS).[10] This approach was subsequently named the qualitative CIOMS method.[9] An improved, mainly quantitative assessment represents the quantitative CIOMS,[9,11,12] which is commonly known as the CIOMS scale.[9]

The MV scale (named for the authors Maria and Victorino) is a purely quantitative method.[13] In the AD method (named for the authors Aithal and Day),[14] causality assessment combines and extends the qualitative CIOMS method,[10] the MV scale,[13] and liver histology results.[14] The ARD method (named for the authors Aithal, Rawlins, and Day)[15] uses, in the first step, some criteria from the qualitative RUCAM[8] and the qualitative CIOMS method,[9,10] and subsequently parts of the AD method,[13] but omits liver histology.[15] The TTK scale (named for the first three authors Takikawa, Takamori, and Kumagi)[16] is a modification of the CIOMS scale.[11] All methods are prospective evaluations, as is the *ad hoc* method (Table 1).[7,28]

The liver-specific method of the Drug Induced Liver Injury Network (DILIN),[17,18] the Causality Assessment Tool (CAT),[19] and the expert opinion method[7,28] are all limited to retrospective evaluation.

Liver-unspecific methods

Various liver-unspecific CAMs also exist[1,20,40] and are still sometimes used to assess liver-related causality in DILI[21] and HILI.[6] Among these are the KL method (named for the authors Karch and Lasagna),[22] the Naranjo scale,[23] and the WHO global introspection method (WHO method).[24] Liver-unspecific CAMs have been used for both prospective and retrospective evaluations (Table 1).

Liver-specific evaluations for prospective use

CAMs suitable for prospective use (Table 1) are of particular clinical importance at the time of clinical presentation, but are also suitable for retrospective evaluation. It is advisable to use an assessment tool that is both prospectively applicable by physicians and retrospectively by the scientific

Table 1. Causality assessment methods for suspected drug-induced and herb-induced liver injury

Causality assessment method	Liver specificity	Prospective evaluation	Retrospective evaluation	Suitability for DILI/HILI
Qualitative RUCAM	+	+		−
Qualitative CIOMS method	+	+		−
CIOMS scale	+	+		+
MV scale	+	+		−
AD method	+	+		−
ARD method	+	+		−
TTK scale	+	+		−
Ad hoc approach	+	+		−
DILIN method	+		+	+
Expert opinion	+		+	+
KL method	−	+		−
Naranjo scale	−	+		−
WHO method	−		+	−

CAMs are specific or unspecific for the liver.[7,28] CAMs primarily based on prospective evaluation are the only tools that allow a prospective strategy for the physician to gather the required items during the disease is ongoing, starting on the day DILI or HILI diagnosis is suspected. Primarily retrospective assessment methods commonly require an expert team, causing delayed evaluation. The CIOMS scale is the preferred tool for prospective assessment by the physician and for retrospective assessment by expert panels, whereas the DILIN method and expert opinion-based assessments are restricted to retrospective evaluation. Details are provided for the following methods:[7] qualitative RUCAM,[8] qualitative CIOMS method,[10] CIOMS scale,[7,9,11,28] MV scale,[13] AD method,[14] ARD method,[15] TTK scale,[16] *ad hoc* approach,[7,28] DILIN method,[17,18] expert opinion,[7,28] KL method,[22] Naranjo scale,[23] and WHO method.[24] Primarily prospectively assessing methods may and should be used retrospectively also.
Abbreviations: AD method, Aithal and Day method; ARD method; Aithal, Rawlins and Day method; CAM, Causality assessment method; CIOMS, Council for International Organizations of Medical Sciences; DILI, Drug-induced liver injury; DILIN, Drug Induced Liver Injury Network; HILI, Herb-induced liver injury; KL method, Karch and Lasagna method; MV scale, Maria and Victorino scale; RUCAM, Roussel Uclaf Causality Assessment Method; TTK scale, Takikawa, Takamori, and Kumagi scale.

community including expert panels, regulatory agencies, and manufacturers.

Qualitative RUCAM

The qualitative RUCAM represented the first objective attempt to assess causality in DILI and considers some characteristic features of liver injury.[8] It uses a qualitative rather than a quantitative approach.[9]

Prospective use

This method does not require an expert group, so it may be used prospectively at the time of suspicion of a liver injury, while the patient is still under treatment by physicians (Table 1).[8] This does not rule out its retrospective application by regulatory agencies, manufacturers, or expert panels.

Liver specificity

The criteria of the qualitative RUCAM are clearly liver-specific (Table 1),[8] although developed from a French method for general drug reaction assessment that was not liver-specific.[25] The original French method was based on chronological and clinical criteria. The chronological criteria included three datasets: time to onset of the reaction, described as very suggestive of, compatible with, or incompatible with drug-induced reaction; the course of the reaction, described as suggestive, non-suggestive, or non-conclusive, which included the clinical course after cessation or continuation of the drug; and response to re-administration, described as positive, negative, or uninterpretable. Responses to these items from the three datasets were combined in a decision table, leading to a chronology score rated as incompatible with, dubious, possible, or suggestive of a drug-induced reaction.

The clinical criteria also included three different items: signs and symptoms suggesting the causal role of the drug and/or presence of a risk factor; result of a specific test proving the causal role of the drug; and assessment of non-drug causes.[8,25] These results were also combined in a decision table, leading to the clinical assessment as dubious, possible, or suggestive.

Finally, chronological and clinical scores were combined, and this resulted in a causality assessment of very likely, likely, dubious, possible, or unlikely.[8,25] Based on chronological and clinical criteria of a general and organ-unrelated assessment, these scores have now been adapted specifically for DILI.[8]

Core elements

The qualitative RUCAM was developed to provide evidence for acute hepatocellular liver injury, which includes a strict definition of liver involvement; precise chronological and clinical criteria suggesting a drug-induced reaction; and a list of tests to exclude other possible causes.[8] Accordingly, acute hepatocellular injury was defined by the highest aminotransferase (AT) activity, so this criterion may apply to either alanine aminotransferase (ALT) or aspartate aminotransferase (AST).[8] However, the minimum AT increase required for the diagnosis was not specified.

Other core elements of the qualitative RUCAM referred to chronological criteria.[8] First, the time to onset of the reaction

was assessed by the dates of the first and last dose of the suspected drug, and a treatment duration of 8–90 days was considered compatible with a suggestive causality, provided the time from the last dose was $\leqslant$ 15 days. A shorter or longer treatment duration was considered compatible, but not suggestive. Second, the course of serum AT activities after cessation of the drug was analyzed. This was very suggestive if the decrease in AT was rapid and reached $\geqslant$ 50% of the difference between the AT peak and the upper limit of normal (N) within 8 days. An AT decrease of $\geqslant$ 50% within 30 days was judged as suggestive for the drug, whereas all other AT changes were either not suggestive or not conclusive. Third, clear basic definitions and conditions were established for the assessment of the response to re-exposure. Required data are the AT levels before re-exposure (designated as baseline AT or ATb) and the AT levels during re-exposure (designated as ATr). Response to re-exposure is measured in multiples of the upper limit of normal as N and is considered positive if ATb is $<$ 5N and ATr $\geqslant$ 2ATb. Other combinations lead to negative or uninterpretable results.

When assessing the clinical criteria for this CAM, signs and symptoms were discussed and considered to be less helpful, as there are no specific drug-induced features.[8] Nevertheless, some risk factors and symptoms, such as fever, rash, and eosinophilia, were mentioned as suggestive of a causative agent. In addition, the lymphocyte transformation test and antibody detection were discussed as evidence for some drugs. Finally, a list of causes unrelated to drugs and a list of necessary tests was compiled. This included hepatitis A, B, and non-A non-B; cytomegalovirus (CMV); Epstein-Barr virus (EBV); herpes virus; alcohol, heart or vascular disease; pregnancy; cancer; and hepatobiliary sonography.

Although the qualitative RUCAM is restricted to acute hepatocellular liver injury,[8] some characteristics of the acute cholestatic and the mixed cholestatic-hepatocellular liver injury were described in a French study in 1987,[26] as explicitly referenced.[8]

Validation

Because of missing reference data, the qualitative RUCAM method could not be validated, and specificity, sensitivity, positive predictive value (PPV) and negative predictive value (NPV) could not be obtained.[8]

Usage frequency

The qualitative RUCAM has not been used in published reports. However, this method was the first approach to specifically assess causality in DILI. The items assessed were vague and qualitative rather than quantitative.[8] This method was, therefore, not suitable for widespread use.[9]

Strengths

The qualitative RUCAM was greatly appreciated as the first preliminary assessment approach for DILI, judging causality ranges based on chronological and clinical criteria.[8]

Weaknesses

Qualitative rather than quantitative item evaluations are characteristic features of this method, which is also limited

to the hepatocellular type of liver injury.[8] The importance of co-medication was not yet properly recognized.

Qualitative CIOMS method

The qualitative CIOMS method[10] represented an improved version of the qualitative RUCAM.[8] It considers the hepatocellular, cholestatic, and mixed cholestatic-hepatocellular types of liver injury,[10] in line with subsequent data.[11]

Prospective use

The qualitative CIOMS method was designed for prospective use by physicians without the need of an expert group, but may be applied retrospectively as well (Table 1).[10]

Liver specificity

For the first time, liver injury was defined, and should be assumed present, if there is an increase of > 2N in ALT or conjugated bilirubin (CB), or if there is a combined increase in AST, alkaline phosphatase (ALP), and total bilirubin (TB), provided one of these is > 2N.[10] No other test result was considered specific for liver disease; in particular, an isolated increase in AST, ALP, or TB even if > 2N should be considered only as a biochemical abnormality, and not necessarily as a sign of liver injury.[10] An increase in ALT, AST, ALP, or TB between N and 2N should be considered as a liver-test abnormality rather than as liver injury.

For the first time, this meant that liver injury was further differentiated by clearly defined criteria.[10] Liver injury is considered hepatocellular if ALT is increased by > 2N alone or R (ratio) is increased ⩾ 5-fold, with R calculated as the ratio of ALT/ALP activity measured together at the time liver injury is suspected, with both activities expressed as multiples of N. Liver injury is considered cholestatic if ALP is increased by > 2N alone or R is ⩽ 2. Liver injury is of the mixed cholestatic-hepatocellular type if both ALT (> 2N) and ALP are increased, and R is > 2 and < 5. Of note, R may vary during the later course of the liver injury.

In studies, acute liver injury required normalization of ALT and ALP within 3 months; otherwise, chronic liver injury was assumed.[10]

Core elements

Core elements of the qualitative CIOMS method for the hepatocellular type of liver injury[10] were similar to or identical with those described for the qualitative RUCAM.[8] However, the suggestive time frame was changed to being 5–90 days from the start of drug administration to onset of the reaction[10] rather than 8–90 days.[8] In addition, rather than using AT to represent ALT or AST,[8] ALT was now the only enzyme used to indicate the reaction and re-exposure test result for the hepatocellular type.[10] Risk factors were expanded to co-medication, and exclusion of drug-unrelated causes should also include hepatitis C virus (HCV), determined by anti-HCV, and alcoholic liver disease, suggested by an AST/ALT ratio of ⩾ 2.[10] Exclusion of CMV and EBV was now optional, and herpes simplex virus (HSV) was no longer considered.

For the first time, core elements for the cholestatic and the mixed cholestatic-hepatocellular type of liver injury were defined by the qualitative CIOMS method.[10] Some core elements differ for the cholestatic and the mixed cholestatic-hepatocellular type[10] compared with the hepatocellular type of liver injury.[8,10] For instance, the time lapse of ⩽ 1 month from drug cessation to the onset of the reaction is considered compatible with causation in cholestatic and mixed liver injury. For the time course after drug withdrawal, it is considered suggestive for the drug if there is > 50% decrease in ALP and/or TB values, expressed as excess over N, occurring within 6 months; the result is considered intermediate if this reduction is < 50% within 6 months.[10] For a positive re-exposure test, a doubling of the ALP is mandatory. To evaluate unrelated causes, ultrasonography of the liver and biliary tract excluding cholelithiasis and biliary tract abnormalities is recommended.

Validation

The qualitative CIOMS method lacks any validation.[10]

Usage frequency

The qualitative CIOMS method has rarely been used in published cases of liver injury, although it was applied in connection with the MV scale as part of the AD method[14] and the ARD method.[15,38]

Strengths

The qualitative CIOMS method[10] extended the qualitative RUCAM,[8] provided a clear definition of the hepatocellular, cholestatic, and mixed cholestatic-hepatocellular types of liver injury, and had some characteristic features.[10] Therefore, a basis for a more stringent case assessment of liver injury was established.

Weaknesses

Assessment by the qualitative CIOMS method was still based mainly on qualitative rather than quantitative scoring of individual items, which weakens its general use.[10] Although the different types of liver injury are clearly defined, with measurements restricted to ALT and ALP only, the general term of 'liver injury' was based on numerous parameters, including ALT, AST, CB, TB, and ALP and, therefore, remained vague.[10] In addition, individual approaches were suggested for the exclusion of alternative causes for different types of liver injury. A uniform approach for all types would have been preferred because the type of liver injury may vary during the later course. Controversy also arose because exclusion of CMV and EBV infections was termed optional, not mandatory, and exclusion of HSV infection was not considered necessary any longer;[10] these recommendations were at variance to the qualitative RUCAM.[8] Consequently, the qualitative CIOMS method should no longer be used.

CIOMS scale

The CIOMS scale was the result of consensus meetings organized at the request of CIOMS[11] and integrated the progress that had been made since the publication of the qualitative RUCAM[8] and the qualitative CIOMS method.[10] The CIOMS scale differs substantially from these other CAMs by being based on quantitatively scored items (Table 2).[11] It is now the most commonly used method

Table 2. Details of the various causality assessment methods for DILI and HILI

Assessed items (with specific scores)	CIOMS	MV	Naranjo	KL	*Ad hoc*	DILIN	WHO	EO
Time frame of latency period (score)	+	+	−	−	−	−	−	−
Time frame of challenge (score)	+	+	−	−	−	−	−	−
Time frame of dechallenge (score)	+	+	−	−	−	−	−	−
Recurrent ALT or ALP increase (score)	+	−	−	−	−	−	−	−
Definition of risk factors (score)	+	−	−	−	−	−	−	−
Verified alternative diagnoses (score)	+	+	−	−	−	−	−	−
Assessed HAV, HBV, HCV (score)	+	+	−	−	−	−	−	−
Assessed CMV, EBV, HSV, VZV (score)	+	+	−	−	−	−	−	−
Liver and biliary tract imaging (score)	+	−	−	−	−	−	−	−
Liver vessel Doppler sonography (score)	+	−	−	−	−	−	−	−
Assessed pre-existing diseases (score)	+	−	−	−	−	−	−	−
Evaluated cardiac hepatopathy (score)	+	−	−	−	−	−	−	−
Excluded alternative diagnoses (score)	+	+	+	−	−	−	−	−
Co-medication (score)	+	−	+	−	−	−	−	−
Prior known hepatotoxicity (score)	+	+	+	−	−	−	−	−
Searched unintended re-exposure (score)	+	+	+	−	−	−	−	−
Defined unintended re-exposure (score)	+	+	−	−	−	−	−	−
Unintended re-exposure (score)	+	+	−	−	−	−	−	−
Laboratory hepatotoxicity criteria	+	+	−	−	−	+	−	+
Laboratory hepatotoxicity pattern	+	+	−	−	−	+	−	+
Liver-specific method	+	+	−	−	−	+	−	+
Structured, liver-specific method	+	+	−	−	−	+	−	−
Quantitative, liver-specific method	+	+	−	−	−	−	−	−
Validated method for hepatotoxicity	+	+	−	−	−	−	−	−

Only items with specific scores were considered, with the exception of the final six assessed items listed in the table. Latency period indicates time from drug/herb initiation to symptoms or abnormal liver tests. The + sign indicates presence and the − sign indicates absence of the items. Data for the DILIN method are derived from the report of Rockey et al.[18] References for the other methods and other details are found in the legend of Table 1.
Abbreviations: ALT, alanine aminotransferase; ALP, alkaline phosphatase; CIOMS, Council for International Organizations of Medical Sciences; CMV, cytomegalovirus; DILI, drug-induced liver injury; DILIN, Drug Induced Liver Injury Network; EBV, epstein-barr virus; EO, expert opinion; HAV, hepatitis A virus; HBV, hepatitis B virus; HCV, hepatitis C virus; HILI, herb-induced liver injury; HSV, herpes simplex virus; KL, Karch and Lasagna method; MV, Maria and Victorino scale; VZV, varicella zoster virus.

for assessing causality in cases of DILI[21] and HILI,[6] both in its original form or its improved and preferred update (Tables 3 and 4).[3–7,21,28]

Prospective use

Physicians treating a patient with liver injury may prospectively use the CIOMS scale to collect the necessary clinical data or to change the diagnostic concept (Table 1).[11] Results are available within a few minutes at the patient's bedside and do not depend upon input from an expert panel.

Liver specificity

The CIOMS scale considers numerous items specific for the liver and liver injury (Table 2).[11] It is a structured scale, and all items for assessment and scoring are quantitative rather than qualitative (Tables 3 and 4).[3–7,9,11,28] Liver injury is defined by an increase in ALT and/or ALP activities of $> 2N$,[11] and there have been recent suggestions to raise the ALT cut-off point to 5N or 3N in the presence of TB values exceeding 2N.[4] Hepatotoxicity is further classified for various types of liver injury: hepatocellular (ALT $> 2N$ alone or R $\geqslant 5$), cholestatic (ALP $> 2N$ alone or R $\leqslant 2$), or mixed cholestatic-hepatocellular

(ALT $> 2N$, increased ALP, with R > 2 and R < 5.[4,7,11,28] This classification is essential because the CIOMS scale differentiates between the hepatocellular (Table 3) and the cholestatic (± hepatocellular) types of liver injury (Table 4).[7,11,28]

Core elements

All core elements of hepatotoxicity (Table 2) are considered in the updated CIOMS scale (Tables 3 and 4): time to onset from beginning or from cessation of the drug/herb intake; course of liver enzyme activities after cessation of the drug/herb; risk factors such as alcohol use, age and pregnancy; co-medication with other drugs/herbs; search for alternative causes; available information on drug/herb hepatotoxicity; and response to unintentional re-exposure.[11] Special emphasis is placed on the results of unintentional re-exposure according to established criteria (Tables 3, 4 and 5).[7,27,28] For the hepatocellular type of injury, the defining criteria are ALT levels before re-exposure (designated as baseline ALT or ALTb), and re-exposure ALT levels (designated as ALTr) (Tables 3 and 5).[7,8,10,11,28] The re-exposure test is positive if ALTb is $< 5N$ and ALTr is $\geqslant 2$ALTb, negative if one or both criteria are not fulfilled, and uninterpretable if data are lacking for one or both criteria. For the cholestatic or the

Table 3. Updated CIOMS scale for the hepatocellular type of injury in DILI and HILI cases

Items for hepatocellular injury	Possible Score	Patient's Score
Time to onset from the beginning of the drug/herb		
5–90 days (rechallenge: 1–15 days)	+2	
< 5 or > 90 days (rechallenge: > 15 days)	+1	
Alternative assessment: Time to onset from cessation of the drug/herb		
⩽ 15 days (except for slowly metabolized chemicals: > 15 days)	+1	
Course of ALT after cessation of the drug/herb		
Percentage difference between ALT peak and N		
Decrease ⩾ 50% within 8 days	+3	
Decrease ⩾ 50% within 30 days	+2	
No information or continued drug/herb use	0	
Decrease ⩾ 50% after day 30	0	
Decrease < 50% after day 30, or recurrent increase	−2	
Risk factors		
Alcohol use (drinks/day: > 2 for women, > 3 for men)	+1	
Alcohol use (drinks/day: ⩽ 2 for women, ⩽ 3 for men)	0	
Age ⩾ 55 years	+1	
Age < 55 years	0	
Concomitant drug(s) or herbs(s)		
None, or no information	0	
Concomitant drug or herb with incompatible time to onset	0	
Concomitant drug or herb with compatible or suggestive time to onset	−1	
Concomitant drug or herb known as hepatotoxin and with compatible or suggestive time to onset	−2	
Concomitant drug or herb with evidence for its role in this case (positive re-challenge or validated test)	−3	
Search for non drug/herb causes		
Group I (6 causes**)**		
Anti-HAV IgM		
HBsAg, anti-HBc IgM, HBV-DNA		
Anti-HCV, HCV-RNA		
Hepatobiliary sonograph /colour Doppler sonography of liver vessels/endosonography/CT/MRC		
Alcoholism (AST/ALT ⩾ 2)		
Acute recent hypotension history (particularly if underlying heart disease present)		
Group II (6 causes)		
Complications of underlying disease(s) such as sepsis, autoimmune hepatitis, chronic hepatitis B or C, primary biliary cirrhosis or sclerosing cholangitis, genetic liver diseases		
Infection suggested by PCR and titer change for CMV (anti-CMV IgM, anti-CMV IgG)		
EBV (anti-EBV IgM, anti-EBV IgG)		
HEV (anti-HEV IgM, anti-HEV IgG)		
HSV (anti-HSV IgM, anti-HSV IgG)		
VZV (anti-VZV IgM, anti-VZV IgG)		
Evaluation of groups I and II		
All causes groups I and II reasonably ruled out	+2	
The 6 causes of group I ruled out	+1	
5 or 4 causes of group I ruled out	0	
< 4 causes of group I ruled out	−2	
Non-drug/herb cause highly probable	−3	
Previous information on hepatotoxicity of the drug/herb		
Reaction labelled in the product characteristics	+2	

Continued

Table 3. Continued

Items for hepatocellular injury	Possible Score	Patient's Score
Reaction published but unlabelled	+1	
Reaction unknown	0	
Response to re-administration		
Doubling of ALT with the drug/herb alone, provided ALT< 5N before re- exposure	+3	
Doubling of ALT with the drug(s) and herb(s) already given at the time of first reaction	+1	
Increase in ALT but < N under the same conditions as for the first administration	−2	
Other situations	0	
Total score for patient		

The compilation of individual items is derived from the updated CIOMS scale,[28] which is based on the original CIOMS scale.[11] The above items specifically refer to the hepatocellular type of injury rather than to the cholestatic (± hepatocellular) type (shown in Table 4). Regarding risk factor of alcohol use, 1 drink commonly contains about 10 g ethanol, and details were discussed recently.[4,30,31] Total score and resulting causality grading: ⩽ 0, excluded; 1–2, unlikely; 3–5, possible; 6–8, probable; ⩾ 9, highly probable.

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CIOMS, Council for International Organizations of Medical Sciences; CMV, cytomegalovirus; CT, computed tomography; DILI, drug-induced liver injury; EBV, Epstein-Barr virus; HAV, hepatitis A virus; HBc, hepatitis B core; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCV, hepatitis C virus; HEV, hepatitis E virus; HILI, herb-induced liver injury; HSV, herpes simplex virus; MRC, magnetic resonance cholangiography; N, upper limit of normal; VZV, varicella zoster virus.

Table 4. Updated CIOMS scale for the cholestatic (± hepatocellular) type of injury in DILI and HILI cases

Items for cholestatic (± hepatocellular) injury	Possible Score	Patient's Score
Time to onset from the beginning of the drug/herb		
5–90 days (rechallenge: 1–90 days)	+2	
< 5 or > 90 days (rechallenge: > 90 days)	+1	
Alternative assessment: Time to onset from cessation of the drug/herb		
⩽ 30 days (except for slowly metabolized chemicals: > 30 days)	+1	
Course of ALP after cessation of the drug/herb		
Percentage difference between ALP peak and N		
Decrease ⩾ 50% within 180 days	+2	
Decrease < 50% within 180 days	+1	
No information, persistence, increase, or continued drug/herb use	0	
Risk factors		
Alcohol use (drinks/day: > 2 for women, > 3 for men) or pregnancy	+1	
Alcohol use (drinks/day: ⩽ 2 for women, ⩽ 3 for men)	0	
Age ⩾ 55 years	+1	
Age < 55 years	0	
Concomitant drug(s) or herbs(s)		
None, or no information	0	
Concomitant drug or herb with incompatible time to onset	0	
Concomitant drug or herb with compatible or suggestive time to onset	−1	
Concomitant drug or herb known as hepatotoxin and with compatible or suggestive time to onset	−2	
Concomitant drug or herb with evidence for its role in this case (positive re-challenge or validated test)	−3	
Search for non drug/herb causes		
Group I (6 causes)		
Anti-HAV IgM		
HBsAg, anti-HBc IgM, HBV DNA		
Anti-HCV, HCV RNA		
Hepatobiliary sonography/colour Doppler sonography of liver vessels/endosonography/CT/MRC		
Alcoholism (AST/ALT ⩾ 2)		
Acute recent hypotension history (particularly if underlying heart disease present)		

Continued

Table 4. Continued

Items for cholestatic (± hepatocellular) injury	Possible Score	Patient's Score
Group II (6 causes)		
Complications of underlying disease(s) such as sepsis, autoimmune hepatitis, chronic hepatitis B or C, primary biliary cirrhosis or sclerosing cholangitis, genetic liver diseases		
Infection suggested by PCR and titer change for CMV (anti-CMV IgM, anti-CMV IgG)		
EBV (anti-EBV IgM, anti-EBV IgG)		
HEV (anti-HEV IgM, anti-HEV IgG)		
HSV (anti-HSV IgM, anti-HSV IgG)		
VZV (anti-VZV IgM, anti-VZV IgG)		
Evaluation of groups I and II		
All causes groups I and II reasonably ruled out	+2	
The 6 causes of group I ruled out	+1	
5 or 4 causes of group I ruled out	0	
< 4 causes of group I ruled out	−2	
Non-drug/herb cause highly probable	−3	
Previous information on hepatotoxicity of the drug/ herb		
Reaction labelled in the product characteristics	+2	
Reaction published but unlabelled	+1	
Reaction unknown	0	
Response to re-administration		
Doubling of ALP with the drug/herb alone, provided ALP < 5N before re-exposure	+3	
Doubling of ALP with the drug(s) and herb(s) already given at the time of first reaction	+1	
Increase in ALP but < N under the same conditions as for the first administration	−2	
Other situations	0	
Total score for patient		

The updated CIOMS scale[28] presented in this table is based on the original CIOMS scale,[11] and was designed specifically for the cholestatic (± hepatocellular) type of liver injury rather than for the hepatocellular type, which differs in a few items and is presented separately in Table 3. Additional details are provided in the legend of Table 3. Total score with resulting causality grading: $\leqslant 0$, excluded; 1–2, unlikely; 3–5, possible; 6–8, probable; $\geqslant 9$, highly probable.
Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CIOMS, Council for International Organizations of Medical Sciences; CMV, cytomegalovirus; CT, computed tomography; DILI, drug-induced liver injury; EBV, epstein-barr virus; HAV, hepatitis A virus; HBc, hepatitis B core; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCV, hepatitis C virus; HEV, hepatitis E virus; HILI, herb-induced liver injury; HSV, herpes simplex virus; MRC, magnetic resonance cholangiography; N, upper limit of the normal range; VZV, varicella zoster virus.

Table 5. Conditions of re-exposure tests in DILI and HILI cases Hepatocellular Cholestatic type of liver injury (± hepatocellular) type of liver injury

Re-exposure test result	ALTb	ALTr	ALPb	ALPr
Positive	< 5N	$\geqslant$ 2ALTb	< 5N	$\geqslant$ 2ALPb
Negative	< 5N	< 2ALTb	< 5N	< 2ALPb
Negative	$\geqslant$ 5N	$\geqslant$ 2ALTb	$\geqslant$ 5N	$\geqslant$ 2ALPb
Negative	$\geqslant$ 5N	< 2ALTb	$\geqslant$ 5N	< 2ALPb
Negative	$\geqslant$ 5N	NA	$\geqslant$ 5N	NA
Uninterpretable	< 5N	NA	< 5N	NA
Uninterpretable	NA	NA	NA	NA

Conditions and criteria for a re-exposure test are described in previous reports.[7,8,10,11,27,28] Required data for the hepatocellular type of liver injury are ALT levels just before re-exposure, designated as baseline ALT (ALTb), and ALT levels after re-exposure, designated as re-exposure ALT (ALTr). Response to re-exposure is positive if both of the following criteria are met: ALTb > 5N and ALTr $\geqslant$ 2ALTb. Other variations lead to negative or uninterpretable test results. For the cholestatic (± hepatocellular) type of liver injury, the corresponding values of ALP are used rather than of ALT.
Abbreviations: ALP, alkaline phosphatase; ALPb, ALP baseline; ALPr, ALP re-exposure; ALT, alanine aminotransferase; ALTb, ALT baseline; ALTr, ALT re-exposure; DILI, drug-induced liver injury; HILI, herb-induced liver injury; N, Upper limit of normal; NA, not available.

mixed cholestatic-hepatocellular injury, the assessment criteria and interpretation of results are similar, with ALT replaced by ALP (Tables 4 and 5).

An update of the original CIOMS scale substantially improved its ability to exclude alternative causes by hepatitis serology, as specific knowledge was gained (Tables 3 and 4).[28] HBsAg and HBV-DNA quantification was added to distinguish hepatitis B virus (HBV) infection from immunization, and HCV-RNA was added to correctly assess HCV infections. In addition, clinical and/or biological parameters for CMV, EBV, or HSV infection had been too vague or were unknown at the time of the initial compilation,[11] and these were specified in the updated CIOMS scale. Infections by hepatitis E virus (HEV) and varicella zoster virus (VZV) were also included and specified (Tables 3 and 4).[28] Specific diagnostic criteria include PCR detection and titer changes of the respective antibodies (IgM, IgG) for CMV, EBV, HEV, HSV, and VZV infections (Tables 3 and 4). The item 'hepato-biliary sonography' was supplemented by color Doppler sonography, including assessments of the liver vessels. Endosonography, computed tomography (CT), and magnetic resonance cholangiography (MRC) were included if these investigations were clinically indicated (Tables 3 and 4). In

recent hepatotoxicity cases, causality has been evaluated by both the updated and the original CIOMS scales, and we found identical results.[27] Therefore, we consider that there is no need for further validation of the updated versus the original CIOMS scale.

Validation

The CIOMS scale was developed by an international expert panel,[11] and validated by cases with known positive re-exposure as gold standard.[12] CIOMS-based assessment has shown good sensitivity (86%), specificity (89%), PPV (93%), and NPV (78%).[12]

Usage frequency

The CIOMS scale in its original or updated form has been widely used for hepatotoxicity assessment in epidemiological studies, clinical trials, case reports, case series, regulatory analyses, and genotyping studies.[7] CIOMS-based results were published by the European Medicines Agency (EMA)[29] and the DILIN group.[17,18] Systematic analyses of CAM usage showed that the original and updated CIOMS scales were the preferred tools in cases of DILI[21] and HILI.[6] Similarly, CIOMS was prioritized by the NIH LiverTox database for causality assessment of hepatotoxicity cases.[30,31]

Strengths

The CIOMS scale is currently the most commonly used tool worldwide to assess causality in hepatotoxicity cases, both prospectively at the time of clinical disease development, and retrospectively by experts. This facilitates the comparability of results because only a single scale is used, rather than a number of different ones. The items are well defined and easily obtained (Tables 3 and 4). In cases of uncertainty, NIH LiverTox provides additional information for some details, as described in its specific search term of causality available from its website,[30,31] as does the international DILI Expert Working Group.[4]

The strengths of the CIOMS scale[11] have been outlined in a number of publications.[3–7,9,28–31] The advantages include stringent criteria for challenge and dechallenge characteristics; exclusion of most relevant alternative causes; assessment of both drugs and herbs; individual evaluation for each co-medicated drug or herb; specific consideration of unintentional re-exposure; unequivocal and liver-specific questions; quantitative individual scores; and a transparent final causality grade, enabled by data transparency and item-by-item data presentation.

We prefer the CIOMS scale over the other CAMs because this scale has a number of advantages as listed in detail (Tables 1 and 2). The CIOMS scale is currently the best CAM for physicians treating a patient with suspected DILI or HILI and can be used to prospectively collect all necessary items without requiring an expert panel. If indicated, subsequent case evaluation may be based on the DILIN method, which allows only retrospective analysis, requires an expert panel, and is so far restricted to the USA.

Weaknesses

The CIOMS scale may be seen as too complex, and an initial causality assessment (pre-test) with a few items derived

from the well-validated CIOMS scale may help to decide whether using this scale is necessary.[9] The pre-test has been used in various types of hepatotoxicity, and the results showed good concordance with the results of the full CIOMS scale.[32–34] Based on qualitative criteria,[11] the pre-test items are intended to establish with only a few questions, whether causality is improbable or not evaluable in hepatocellular or cholestatic (± hepatocellular) injury.[9,32–34]

Some refinement and strengthening of the CIOMS scale is recommended for items such as alcohol, age, and pregnancy.[3,4,35] To extend the short list of alternative diagnoses in the CIOMS scale (Tables 3 and 4), and to consider rare causes unrelated to drugs or herbs, improved checklists are available[7,9,28] to be used as post-test tools after the CIOMS scale is used.[7,9,28]

MV scale

The MV scale[13] is similar to, but shorter than the CIOMS scale.[11] It was developed in an attempt to improve upon the CIOMS scale by deleting some items, adding clinical elements, and simplifying and changing the relative weight of elements in the assessment of causality.[9,30,31]

Prospective use

The MV scale is suitable for prospective use by physicians and does not require an expert panel (Table 1).[13]

Liver specificity

The items of the MV scale are liver-specific.[9,13,28]

Core elements

The MV scale consists of five core elements: temporal relationship, exclusion of alternative causes, extrahepatic features, re-exposure test, and previous hepatotoxicity reports.[13] Selection criteria were based on the personal experience of the two authors, and on the medical literature. The relative weight of each component was analyzed, and component scores were attributed.[13] There are major differences between the MV scale and the CIOMS scale regarding quality, quantity, and scoring of individual elements (Table 2).[3,9,11,28,30,31]

Validation

Data on specificity, sensitivity, PPV, and NPV for the MV scale are not available.[13] The method was assessed for content, construct, criterion, and inter-rater reliability, with varying results.

Usage frequency

The MV scale was used in three DILI studies as part of the AD and ARD methods,[14,15,38] but not in 38 other publications of DILI cases[21] or in 23 publications of HILI cases.[6]

Strengths

The MV scale is a liver-specific, structured, and quantitative causality algorithm providing scores and different levels of causality.[13] This scale has performed well in tests comparing

it with expert opinion.[13,30,31,36] Compared to the CIOMS scale,[11] the MV scale was equivalent in cases of drug hypersensitivity; for other etiologies, the CIOMS scale appeared superior,[35,36] and was generally considered more reliable.[3]

Weaknesses

Compared with the original CIOMS scale,[11] the MV scale[13] has a number of shortcomings and lacks equivalency, which are of concern (Table 2).[3–5,9,28,30,31,35–37] A comparison of MV scale and CIOMS scale results was disappointing overall[35] because of low consistency between results;[35–37] complete agreement between the scales was present in only 18% of cases.[3,36] The CIOMS scale showed better discriminative power, and its assessments were also closer to those of specialists.[36] A recent HILI study confirmed poor concordance between the MV and CIOMS scales for both the herb and concomitant medication assessments,[37] with higher causality levels for both assessments given by the CIOMS scale compared with the MV scale. The low MV scores were attributed to different considerations of prolonged latency and dechallenge periods; the presence of several alternative, herb-independent causes for the observed liver disease; only partial exclusion of herb-unrelated causes, due to missing essential case data; and lack of consideration of extrahepatic features such as rash, fever, arthralgia, peripheral eosinophilia, and cytopenia. It, therefore, appears that various confounders prevent the MV scale from identifying a high level of causality for a particular herb when assessing HILI cases.

The MV scale has fewer liver-specific criteria[13] than the original CIOMS scale,[11] evaluates dechallenge as the time necessary for ALT or ALP to fall to < 2N, and considers a shorter latency period.[13] Therefore, it performs poorly in atypical cases such as those with unusually long latency periods or residual chronic features after cessation.[3,30,31,36] This scale is less accurate than the CIOMS scale for exclusion of drug-independent causes, ignores concomitant drug use, underestimates drugs that have been marketed for > 5 years without published cases of hepatotoxicity, and overestimates extrahepatic features such as hypersensitivity reactions,[3,13,30,31] considering that hypersensitivity reactions are comparatively infrequent in hepatotoxicity cases.[3,36]

Validation of the MV scale used real and fictitious cases and the opinion of three external experts; there was agreement between the scale and experts in 84% of cases.[3,13] Concern remained that the authors' estimation of the scale assessment weight might have enhanced the scale's performance,[3] as it did not use cases with verified positive re-exposure tests[13] and, therefore, differed from the approach of the CIOMS scale, which used both positive re-exposure tests and a panel of experts.[11,12] NIH LiverTox specifically criticized the fact that the identification of elements and their relative weights in the MV scale were based on the expert opinion of the two authors, rather than by prospective evaluation using different elements, or by modeling different cut-off points and weights.[30,31] In addition, the low numbers of experts and the low degree of validation of the MV scale[13] was criticized.[30,31] Because of these limitations, the MV scale is not commonly recommended for assessing causality in assumed DILI and HILI cases, and is certainly no substitute for the CIOMS scale.[3,28,30,31,35–37]

AD method

The AD method[14] is not an independent method, but rather represents a combination of the qualitative CIOMS method,[10] the MV scale,[13] and an index liver histology.[14]

Prospective use

In principle, the AD method is suitable for prospective use (Table 1);[14] however, liver histology results are not commonly available or only later in the assessment course.

Liver specificity

Through its individual components, the AD method is liver-specific (Table 1).[14]

Core elements

The AD method requires an index liver biopsy to rule out an alternative cause of the liver disease.[14] Assessment by some items of the qualitative CIOMS method[10] is used to grade any reaction; if this indicates a possible drug-related liver injury, the MV grade of causality is assumed if the score is ⩾ 11 points.[9,14] However, the original MV scale defines a score of 10–13 points as possible causality only.[13]

Validation

The AD method is not validated.[9,14]

Usage frequency

The AD method as described[14] has not been used in other published studies.

Strengths

There has been no convincing evidence for the strengths of this particular method.[14]

Weaknesses

The AD method[14] combines the weaknesses of the qualitative CIOMS method[10] and the MV scale.[13] In addition, liver histology does not specifically support or exclude causality in hepatotoxicity cases, and thus is not commonly recommended.[7] It remains unclear why elements of the unvalidated qualitative CIOMS method[10] rather than the more appropriate CIOMS scale[11] were incorporated into the AD method.[14] Critical reports of the validity of the MV scale appeared in 2001[35,36] and were not available at the time when the AD method was published in 1999.[14] Overall, the AD method is complex, hampers objective evaluations, and is not recommended for common use.[9]

ARD method

The ARD method again is not an original approach.[15] It is based on the results of consensus meetings, referring to the qualitative RUCAM,[8] the qualitative CIOMS method,[10] and the AD method.[14] Liver histology is not required; however, it is unclear whether the MV scale was used as part of the AD method.[15] An updated version of the ARD method[38] was

based on the qualitative CIOMS method[10] and the AD method,[14] including the MV scale;[13] again, liver histology was not required.[38]

Prospective use

Prospective use is possible for the ARD method (Table 1),[15] including its update.[9,38]

Liver specificity

Because of the components used, both the ARD method[15] and its update[38] are liver-specific (Table 1).

Core elements

The core elements of the ARD method correspond to those of its individual components.[15,38]

Validation

No validation has been published for the original ARD method.[15] For its update, data for sensitivity, specificity, and PPV were presented for cases with indeterminate causality.[38] However, this particular group of cases has basic assessment problems.[38,41] In particular, for cases of drug-related liver injury with inconclusive causality, the sensitivity of the updated ARD method increased, but its specificity decreased,[38] creating some concern.[41]

Usage frequency

The ARD method and its update are not commonly used in hepatotoxicity cases.[6,21]

Strengths

Evidence for the strengths of the ARD method and its update is lacking.[15,38,41] Initial core elements included items of the qualitative CIOMS method, allowing classification of a liver injury as drug-related, drug-unrelated, or intermediate.[15,38] Subsequent core elements were derived from the MV scale.[38]

Weaknesses

Among the key problems of the ARD method[15] and its update[38] is their use of the qualitative RUCAM,[8] the qualitative CIOMS method,[10] or possibly the MV scale[13] as part of the AD method,[14] rather than the CIOMS scale itself.[15,38] The weaknesses of the qualitative CIOMS method and/or the MV scale have been summarized above and previously.[3,9,41] They were carried over when used as components of the ARD method.[15,38] The ARD method is complex, disputed, and cannot be recommended for general use.

TTK scale

The TTK scale[16] represents a major modification of the CIOMS scale,[11] as shown in a recent tabulated compilation.[3] These modifications include a greater emphasis on drug reactions triggering immunological responses such as inclusion of the drug lymphocyte stimulation test (DLST).[16]

Prospective use

Similar to the CIOMS scale,[11] the TTK scale is suitable for prospective use (Table 1).[16]

Liver specificity

The scale contains items that are liver-specific (Table 1).[3,16]

Core elements

Compared with the CIOMS scale,[11,12] major modifications are evident in the TTK scale.[16] These include different evaluations of the chronological data, exclusion of co-medication, implementation of the DLST, and inclusion of eosinophilia into the assessment system.[16] The TTK scale elements also differ substantially from those of the MV scale.[3]

Validation

Validation of the TTK scale is incomplete as the sensitivity, specificity, PPV, and NPV have not been assessed.[16,39] In one study, sensitivity values for the TTK, CIOMS, and MV scales were 93.8%, 77.8%, and 43.2%, respectively; the corresponding values for specificity were 89.1%, 100%, and 100%, respectively.[42] The weighted statistical test indicated a poor correlation between the results from the TTK and the CIOMS scales.[42]

Usage frequency

The TTK scale is widely used in Japan,[16] and has been recently reviewed.[39] In other countries, this scale is not or is only rarely considered for use.[3–6,9,20,21,30,31] Limited access and lack of standardization have prevented generalized clinical use of the DLST and consequently of the TTK scale outside Japan;[3] this may be due to methodological difficulties with false-positive and false-negative cases in the DLST.[16]

Strengths

Compared with the CIOMS scale[11] and the MV scale,[13] the TTK scale may be superior in Japanese cases.[3,16,39,42] Despite active contribution from Japan, the international DILI Expert Working Group did not consider the proposals made in the TTK scale,[16] nor did NIH LiverTox.[30,31]

Weaknesses

It remains to be established whether the TTK scale is superior to other CAMs, as this scale selectively includes and excludes core elements, thereby possibly facilitating a high total score. Initially higher causality levels were cited as evidence for superiority of the TTK scale over the CIOMS scale.[16] However, differences between these scales in individual items,[3,9,16,39,42] scoring values of items,[3,16,39,42] and ranges for the final scores[3] resulted in discrepancies in the final scores obtained by different authors using the TTK scale.[3,42] With the TTK scale, a higher causality level is easily achieved through the addition of DLST and eosinophilia and exclusion of an obligatory co-medication assessment, which downgrades causality in other scoring systems.[16]

In a Japanese study based on parameter variations,[42] the TTK scale was considered possibly superior to both the CIOMS

and the MV scales in the diagnosis of DILI.[3,42] This was explained by the finding that the distribution of cases into probability categories by the TTK scale results in higher probability rates than those given by the CIOMS and the MV scales.[42] However, the proposed superiority is unwarranted because core elements of the TTK scale may be added or subtracted selectively, leading to erroneously high causality gradings.[16] It has also been suggested that the TTK scale is able to diagnose DILI more accurately,[3,42] as shown by cases that have been assessed as being without causality using the CIOMS scale.[42] Indeed, patients with liver disease such as EBV, HAV and HCV infections, hepatocellular carcinoma, acute circulatory failure, and drug use (including over-the-counter drugs) were classified as non-DILI cases by the CIOMS scale and as DILI cases by the TTK scale.[42] Against this is the possibility that the TTK scale may over-diagnose and over-report DILI cases, as DILI and HILI are diagnoses of exclusion. In addition, receiver operating characteristic curves could not establish evidence for superiority of the TTK scale; these curves revealed only that both the CIOMS and the TTK scales are probably superior to the MV scale in terms of discrimination,[42] confirming other studies.[3] Thus, the TTK scale presently is not a preferred tool.

Ad hoc method

The *ad hoc* method is used prospectively as soon as DILI or HILI is suspected by physicians familiar with hepatotoxicity, but not necessarily with sophisticated CAMs. It has also been used in publications related to DILI[21] and HILI.[6]

Prospective use

Prospective use of this method is common while the patient is being treated by the physicians experienced in hepatotoxicity (Table 1).

Liver specificity

In patients with suspected hepatotoxicity, liver-specific criteria are considered globally, but not defined in detail.[7,28,35,43]

Core elements

Although proposed items such as symptoms, disease signature, latency period, dechallenge, definitive exclusion of alternative causes, risk factors, alcohol use, and product track record are in use, no universally accepted description exists for this method or its application.[7,28,35,43]

Validation

The *ad hoc* method is not validated.

Usage frequency

Published DILI and HILI reports lacking any description of CAM are presumably based on the *ad hoc* method. This applies to 38 of 61 DILI publications (62%)[21] and to 3 of 23 HILI publications (13%).[6] NIH LiverTox does not explicitly mention the *ad hoc* approach as a CAM for hepatotoxicity cases.[30,31]

Strengths

There are no obvious strengths over other approaches (Table 2).

Weaknesses

Initial use of the *ad hoc* assessment[7,28,35,43] prior to the liver-specific CIOMS scale[11] will inevitably delay the final and valid assessment, and increase the number of missed alternative diagnoses commonly described in initially suspected DILI[7,14,15] and HILI.[6,7] Lack of validation and transparency renders the *ad hoc* approach obsolete for assessment of causality in suspected DILI and HILI cases.

Liver-specific evaluations for retrospective use

Methods for retrospective causality analysis of DILI and HILI cases (Table 1) are of little clinical relevance to physicians in need of early results when therapeutic decisions have to be made.

DILIN method

According to NIH LiverTox, the DILIN method was compiled by analysis of a condensed narrative summary, a summary of clinical findings, and sequential biochemical abnormalities,[30,31] extracted from clinical records and entered into a 65-page case report form.[18] The DILIN causality adjunction process is delineated in a 12-step flow diagram for three independently assessing experts in hepatotoxicity, who grade the likelihood of a causal relationship between the drug and liver injury by one of five scores.[18] NIH LiverTox briefly mentions the DILIN method,[30,31] as have others.[3,4]

Another approach of the DILIN group uses a novel CAT specifically for herbs and dietary supplements (HDS), which was presented as an abstract.[19] In this preliminary study, CAT was used for 16 DILI cases initially evaluated by the DILIN method, and HDS were implicated as a potential cause.

Retrospective use

The DILIN method is to be used retrospectively (Table 1).[17,18] In addition, using structured causality assessment and expert opinion, CAT was designed to retrospectively adjudicate multiple products as a single entity.[19]

Liver specificity

The items of both the DILIN method (Table 1)[17,18] and the CAT[19] are liver-specific.

Core elements

To retrospectively exclude alternative causes, the DILIN method screens for previous liver disease, alcohol use, serological and virological evidence of hepatitis A, B, or C infection, autoantibodies, ceruloplasmin, α-1-antitrypsin, ferritin, iron, and imaging data; however, no specific details or appropriate scores for each item were provided (Table 2).[18]

The CAT elements include multiple items of HDS products, implicated drugs, alternative diagnoses, and published cases of adverse reactions related to the product or its

ingredients.[19] Analogous to the scoring system of the DILIN method, which expresses causality levels as percentage assurance,[18,28,30,31] CAT also grades the likelihood of a causal relationship between HDS and liver injury from definitive to unlikely.[95]

Validation

Validation included the level of complete agreement between the reviewers, which was reported as 27% with the DILIN method versus 19% with the CIOMS scale, and the two scales had a modest correlation with each other.[18] In addition, the CIOMS scale was more conservative and substantially shifted the causality likelihood toward the lower probabilities compared with the DILIN method. In the CAT study, overall agreement and reliability was moderate.[19] This method needs further investigation and validation.[5]

Usage frequency

The DILIN method was used in 4.3% of 23 publications of cases initially suspected as HILI[6,44] as well as in various DILI studies, although there are fewer reports for the latter.[17,18,45–47]

Strengths

The DILIN method attempts to resolve the complexity of hepatotoxicity causality assessment by a complex, retrospective evaluation,[18] as does CAT.[19] The DILIN method, especially when combined with the CIOMS scale, may well be suited for retrospective studies,[18] and could be the basis for future valid studies of host, genetic, environmental, and immunological risk factors to be carried out by the DILIN group.[45,46]

Weaknesses

The DILIN method requires experts,[17,18,45–47] and was used for retrospective assessments of case series when time to conclusion is not a crucial issue.[7,28] It is, therefore, not suitable for prospective use at the beginning of a disease. The method is complex and needs multiple steps, including completion of a 65-page case report form.[18] Although alternative infectious causes such as HEV, CMV, EBV, HSV, and VZV are commonly assessed in careful analyses of initially assumed DILI and HILI cases[6,7,15,27,28,42,47] and are components of the updated CIOMS scale (Tables 3 and 4),[28] these causes are ignored by the DILIN assessment method.[18] Neglecting clinically important alternative causes may partially explain why high likelihood scores obtained with the DILIN method are shifted to lower scores when the CIOMS scale is used.[18] When HEV infections were overlooked in cases initially assumed to be DILI, which were evaluated by the DILIN method,[47] it appeared that the DILIN method is at risk of over-diagnosing and over-reporting DILI and HILI. Preference should be given to lower case numbers with thorough causality evaluation rather than to high case numbers achieved by less stringent assessment methods.

The DILIN method is used mainly in the USA, and has not found wider acceptance. Transparency of causality results obtained with the DILIN method is low,[18] but transparent data and results are preferable to a simple final causality grading. Item-by-item data presentation is also feasible with the updated CIOMS scale (Tables 3 and 4), as shown for a few examples.[6,27,28,32–34,37,48]

Expert opinion method

CAMs based on expert opinion or expert panels are poorly defined, requiring specialists with clinical expertise in hepatology to be available for causality assessment in DILI and HILI,[30,31] as detailed previously.[27]

Retrospective use

Assessment is retrospective.[27,30,31]

Liver specificity

Core elements are not commonly described, unless in the context of a specific causality assessment by an expert panel.

Validation

Depending on the individual approach, results of validation may be available, but have not been published.

Usage frequency

Because the expert opinion approach is not defined, no valid data for its use are available.

Strengths

For DILI assessment, skilled hepatologists are available in most countries including Japan,[16,39,42] especially in expert panels such as the international DILI Expert Working Group,[4] the DILIN group,[17,18,45–47] the Spanish Group for the Study of Drug-Induced Liver Disease,[3,49] and the Spanish-Latin American network on DILI.[50] For HILI, the Hong Kong Herb-Induced Liver Injury Network (HK-HILIN) is of importance,[51] as are other groups.[19,44,52]

Weaknesses

Qualification of assessors is crucial and may be a problem, as discussed recently.[53–55] Even with specialists, individual opinion often results in judgment differences.

Liver-unspecific causality assessment methods

For DILI and HILI cases, liver-unspecific CAMs are obsolete (Table 1).[3,5–7,27,28,37,43] However, as some methods have been used in the past, and are briefly discussed below.

KL method

The KL method[22] is neither liver-specific (Tables 1) nor validated for hepatotoxicity, and lacks important items for hepatotoxicity (Table 2), as discussed recently.[27,28] It has been used for causality assessment of suspected herbal hepatotoxicity.[56] Subjective judgment is needed in many steps, making this method more prone to bias.[3] Although in common use by the Spanish Pharmacovigilance Centres,[56] the KL method is not used by the Spanish Group for the Study of Drug-Induced Liver Disease,[3,49,52] which exclusively utilizes the CIOMS scale as the preferred assessment tool.

The KL method should not be used for assessment of hepatotoxicity cases.[27,28]

Naranjo scale

The use of the Naranjo scale[23] in hepatotoxicity cases is problematic,[3,6,7,30,31,43,57–59] as detailed recently.[28,53] This scale is liver-unspecific (Tables 1 and 2) and was designed to assess causality for any ADR independent of the affected organ.[23] It relates toxic drug reactions to general pharmacological drug actions, and thus has a lower sensitivity for rare and idiosyncratic reactions such as those prevalent in liver injury.[23] The scale considers drug concentrations and monitoring, dose relationship including decreasing dose, placebo response, cross-reactivity, and confirmation of ADRs using unidentified objective evidence, which is irrelevant for DILI and HILI.[3,6,7,28,30,31,43,53,57–59] In essence, the Naranjo scale is obsolete in causality assessment of DILI and HILI cases.

WHO method

The WHO method[24] is not liver-specific, was not developed or validated for hepatotoxicity cases, and does not consider hepatotoxicity-related characteristics (Tables 1 and 2).[6,7,27,28,43] These shortcomings have raised major concerns[7,28,54,55] and led to the conclusion that this scale is neither appropriate for causality assessment in suspected hepatotoxicity cases[7,55,60,61] nor has advantages over other causality algorithms.[7,28] The WHO method was not specifically mentioned, addressed, or discussed as a CAM for hepatotoxicity cases in relevant reports,[3–5,9,35] including a recent statement from NIH LiverTox.[30,31] This method is obsolete for hepatotoxicity case assessment.[7,28,43,55,60–62]

Considerations for future strategies

Over the past decades, substantial progress has been made in DILI and HILI research, and various international consensus meetings have established liver-specific CAMs related to DILI and HILI cases. The qualitative RUCAM[8] and the qualitative CIOMS method[10] were important preliminary liver-specific tools, and valuable precursors to the quantitative, structured and liver-specific CIOMS scale,[11] which was validated[12] and updated (Tables 3 and 4).[9,28] In various hepatotoxicity studies, causality was assessed by both the updated and the original CIOMS scale, and identical results were obtained, substantiating validation of the updated CIOMS scale compared with the original CIOMS scale. Therefore, the updated scale did not require re-validation.[27,58,60,61,63,64] The CIOMS scale and its update are now commonly used tools worldwide for assessment of causality in DILI and HILI cases.

In the future, stringent efforts will be needed to ensure continuous use and further improvement of the CIOMS scale in its updated form by physicians treating patients with DILI and HILI. The prospective approach will improve quality of case data, validity of causality assessment, and clinical outcome by reducing the risk of missed diagnoses. On the day that DILI or HILI is suspected, a CIOMS-based, item-by-item-causality assessment should be initiated (Tables 3 and 4). This ensures early estimation of the likely causality level and facilitates prospective completion of the data collection by the itemized CIOMS list, as shown in a recent case report

of severe hepatotoxicity caused by Indian Ayurvedic herbal products.[28,32] An exhaustive checklist for alternative causes is available and should be used as a reminder to exclude or to establish other diagnoses unrelated to DILI and HILI, avoiding missed diagnoses.[7,28] When probable or highly probable causality is established, the case may be diagnosed as DILI or HILI based on the completed CIOMS scale and the checklist. This case report represents the collection and presentation of all raw data including sequential biochemical abnormalities, along with a summarizing narrative case report, facilitating the follow-up. The collected data may be presented in anonymized form to the scientific community, other expert panels, regulatory agencies, and manufacturers for further evaluation if needed.

Collection of data, including the individual CIOMS items, will serve as a basis for retrospective re-evaluation of prospectively collected data of excellent quality. Thus, requests for further data and expert discussions will be replaced by stringent case evaluation and will obviate the need for additional efforts such as the retrospective use of the DILIN method. This retrospective method with expert-based analyses shows major inter-rater problems, is rather complex to use, and lacks transparency of causality assessment results for individual cases.[18,30,31] Good and reproducible causality assessment needs excellent data from the beginning of the DILI and HILI disease, with transparent case data and causality assessment details. This is preferred over questionable attempts to compensate for earlier shortcomings in data collection and evaluation and/or inter-rater concordance problems related to questionable quality of case data. Indeed, published reports often do not provide the data needed to determine hepatotoxicity causality in initially suspected cases of DILI, as shown by the DILIN group,[65] or in HILI, as reported by others.[6,7,27,28,33,34,58,60,61,63,64] Input of good-quality data into a valid system should lead to output of good results, whereas poor results are frequently a consequence of poor-quality data input. Thus, early and appropriate data collection and evaluation are the key issues, rather than attempts at subsequent compensation and correction.

Future reports of DILI and HILI cases should ensure full transparency of complete case data, including the tabulated CIOMS scale for the individual patient, as shown previously for hepatotoxicity cases of single case reports,[28,32,48] case series,[27,33,34,37] and spontaneous reports to regulatory agencies.[58,60,61] Inclusion of listed essential diagnostic elements in research articles could increase the quality and clinical utility of hepatotoxicity case reports, in line with suggestions made by the DILIN group.[65] To prevent a flood of cases with unsubstantiated causality, publication should be limited to cases with a probable or highly probable CIOMS causality level. Future efforts should be directed at dismissing obsolete CAMs for DILI and HILI, that is, methods that are not liver-specific.

Additionally, assessing causality in DILI and HILI cases should follow a pragmatic strategy, identical in all countries, to allow comparability and international harmonization. On the day of suspicion, causality evaluation should start with the collection of all necessary data and use of the CIOMS scale, in line with proposals made recently by the international DILI Expert Working Group from Europe, the USA, and Japan.[4] This standardized approach should improve validity of causality assessments in DILI and HILI cases.

Conflict of interest

None

Author contributions

Substantial contributions to the conception and design (RT, JS), literature search and acquisition of relevant literature (AE, JS), analysis and interpretation of the data (RT, AE, JS), drafting the article (RT), critical revision of important intellectual content (AE, JS), final approval of the version to be published (RT, AE, JS).

References

[1] Aronson JK. Adverse drug reactions: History, terminology, classification, causality, frequency, preventability. In: Talbot J, Aronson JK, ed. Stephens' Detection and Evaluation of Adverse Drug Reactions: Principles and Practice. 6th Edition. New York: John Wiley & Sons, Ltd, 2011:1–119.

[2] Aronson JK. Meyler's side effects of herbal medicines. Amsterdam: Elsevier,. 2009.

[3] García-Cortés M, Stephens C, Lucena MI, Fernández-Castañer A, Andrade RJ. Causality assessment methods in drug induced liver injury: Strengths and weaknesses. J Hepatol 2011;55:683–691.

[4] Aithal GP, Watkins PB, Andrade RJ, Larrey D, Molokhia M, Takikawa H, et al. Case definition and phenotype standardization in drug-induced liver injury. Clin Pharmacol Ther 2011;89:806–815.

[5] Bunchorntavakul C, Reddy KR. Herbal and dietary supplement hepatotoxicity. Aliment Pharmacol Ther 2013;37:3–17.

[6] Teschke R, Schulze J, Schwarzenboeck A, Eickhoff A, Frenzel C. Herbal hepatotoxicity: suspected cases assessed for alternative causes. Eur J Gastroenterol Hepatol 2013; in press.

[7] Teschke R, Schwarzenboeck A, Eickhoff A, Frenzel C, Wolff A, Schulze J. Clinical and causality assessment in herbal hepatotoxicity. Expert Opin Drug Saf 2013;12:339–366.

[8] Danan G. Consensus meetings on: causality assessment of drug-induced liver injury. J Hepatol 1988;7:132–136.

[9] Teschke R, Schwarzenboeck A, Hennermann KH. Causality assessment in hepatotoxicity by drugs and dietary supplements. Br J Clin Pharmacol 2008; 66:758–766.

[10] Bénichou C. Criteria of drug-induced liver disorders. Report of an international consensus meeting. J Hepatol 1990;11:272–276.

[11] Danan G, Bénichou C. Causality assessment of adverse reactions to drugs – I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. J Clin Epidemiol 1993; 46:1323–1330.

[12] Bénichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs – II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. J Clin Epidemiol 1993;46: 1331–1336.

[13] Maria VA, Victorino RM. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. Hepatology 1997;26:664–669.

[14] Aithal PG, Day CP. The natural history of histologically proved drug induced liver disease. Gut 1999;44:731–735.

[15] Aithal GP, Rawlins MD, Day CP. Accuracy of hepatic adverse drug reaction reporting in one English health region. BMJ 1999;319:1541.

[16] Takikawa H, Takamori Y, Kumagi T, Onji M, Watanabe M, Shibuya A, et al. Assessment of 287 Japanese cases of drug induced liver injury by the diagnostic scale of the International Consensus Meeting. Hepatol Res 2003; 27:192–195.

[17] Rochon J, Protiva P, Seeff LB, Fontana RJ, Liangpunsakul S, Watkins PB, et al. Reliability of the Roussel Uclaf Causality Assessment Method for assessing causality in drug-induced liver injury. Hepatology 2008;48:1175–1183.

[18] Rockey DC, Seeff LB, Rochon J, Freston J, Chalasani N, Bonachini M, et al. Causality assessment in drug-induced liver injury using a structured expert opinion process: Comparison to the Roussel-Uclaf causality assessment method. Hepatology 2010;51:2117–2126

[19] Navarro VJ, Barnhart HX, Bonkovsky HL, Reddy KR, Seeff L, Serrano J, et al. Diagnosing hepatotoxicity attributable to herbal and dietary supplements: test-retest reliability of novel causality assessment tool. J Hepatol 2012;56 (Suppl 2):S536.

[20] Agbabiaka TB, Savovic J, Ernst E. Methods for causality assessment of adverse drug reactions: A systematic review. Drug Saf 2008;31:21–37.

[21] Tajiri K, Shimizu Y. Practical guideline for diagnosis and early management of drug-induced liver injury. World J Gastroenterol 2008;14:6774–6785.

[22] Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. Clin Pharmacol Ther 1977;21:247–254.

[23] Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther 1981;30:239–245.

[24] WHO, World Health Organization. The use of the WHO-UMC system for standardised case causality assessment. WHO Collaborating Centre for International Drug Monitoring (Uppsala Monitoring Centre, UMC), Database 2000. Available at: http://who-umc.org/Graphics/24734.pdf Accessed 30 March 2013.

[25] Bégaud B, Evreux JC, Jouglard J, Lagier G. Unexpected or toxic drug reaction assessment (imputation). Actualization of the method used in France. Therapie 1985;40: 111–118. Article in French.

[26] Danan G, Bénichou C, Begaud B, Biour M, Couzigou P, Evreux JC, et al. Criteria of causality of an acute hepatitis by drugs. Results of consensus meetings. Gastroenterol Clin Biol 1987;11:581–585.

[27] Teschke R, Frenzel C, Schulze J, Schwarzenboeck A, Eickhoff A. Herbalife hepatotoxicity: Evaluation of cases with positive reexposure tests. World J Hepatol 2013;5:353–363.

[28] Teschke R, Frenzel C, Schulze J, Eickhoff A. Herbal hepatotoxicity: challenges and pitfalls of causality assessment methods. World J Gastroenterol 2013; 19:2864–2882.

[29] EMA (European Medicines Agency): Assessment of case reports connected to herbal medicinal products containing cimicifugae racemosae rhizoma (black cohosh, root). Issued May 8, 2007. Available at: http://www.ema.europa. eu/docs/en_GB/document_library/Herbal_-_HMPC_assessment_report/ 2010/02/WC500074167.pdf Accessed 30 March 2013.

[30] National Institutes of Health (NIH): NIH launches free database of drugs associated with liver injury, October 12, 2012 News Release. Available at: http://www.nih.gov/news/health/oct2012/niddk-12.htm Accessed 30 March 2013.

[31] National Institutes of Health (NIH) and LiverTox: Drug record. Herbals and dietary supplements. Last updated 20 February 2012. Available at: http:// www.livertox.nih.gov/Herbals_and_Dietary_Supplements.htm Accessed 30 March 2013.

[32] Teschke R, Bahre R. Severe hepatotoxicity by Indian Ayurvedic herbal products: A structured causality assessment. Ann Hepatol 2009;8:258–266.

[33] Teschke R, Bahre R, Fuchs J, Wolff A. Black cohosh hepatotoxicity: quantitative causality evaluation in nine suspected cases. Menopause 2009;16:956–965

[34] Teschke R, Schwarzenboeck A. Suspected hepatotoxicity by cimicifugae racemosae rhizoma (black cohosh, root): critical analysis and structured causality assessment. Phytomedicine 2009;16:72–84.

[35] Kaplowitz N. Causality assessment versus guilt-by-association in drug hepatotoxicity 2001;33:308–310.

[36] Lucena MI, Camargo R, Andrade RJ, Perez-Sanchez CJ, Sanches de la Cuesta F. Comparison of two clinical scales for causality assessment in hepatotoxicity. Hepatology 2001;33:123–130.

[37] Teschke R, Fuchs J, Bahre R, Genthner A, Wolff A. Kava hepatotoxicity: comparative study of two structured quantitative methods for causality assessment. J Clin Pharm Ther 2010;35:545–563

[38] Aithal GP, Rawlins MD, Day CP. Clinical diagnostic scale: a useful tool in the evaluation of suspected hepatotoxic adverse drug reactions. J Hepatol 2000; 33:949–952.

[39] Takikawa H. Recent status of drug-induced liver injury and its problems in Japan. Jap Med Ass J 2010;53:243–247.

[40] Macedo AF, Marques FB, Ribeiro CF, Teixeira F. Causality assessment of adverse drug reactions: comparison of the results obtained from published decisional algorithms and from the evaluations of an expert panel, according to different levels of imputability. J Clin Pharm Ther 2003;28:137–143.

[41] Lee WM. Assessing causality in drug-induced liver injury. J Hepatology 2000; 33: 1003–1005.

[42] Watanabe M, Shibuya A. Validity study of a new diagnostic scale for drug-induced liver injury in Japan – comparison with two previous scales. Hepatol Res 2004;30:148–154.

[43] Teschke R, Wolff A. Regulatory causality evaluation methods applied in kava hepatotoxicity: Are they appropriate? Regul Toxicol Pharmacol 2011;59:1–7.

[44] Fong TL, Klontz KC, Canas-Coto A, Casper SJ, Durazo FA, Davern TJ, et al. Hepatotoxicity due to Hydroxycut: A case series. Am J Gastroenterol 2010; 105:1561–1566.

[45] Fontana RJ, Watkins PB, Bonkovsky HL, Chalasani N, Davern T, Serrano J, et al. Drug-induced liver injury Network (DILIN) prospective study: rationale, design and conduct. Drug Saf 2009;32:55–68.

[46] Chalasani N, Fontana RJ, Bonkovsky HL, Watkins PB, Davern T, Serrano J, et al. Causes, clinical features, and outcomes from a prospective study of drug-induced liver injury in the United States. Gastroenterology 2008;135: 1924–1934.

[47] Davern TJ, Chalasani N, Fontana RJ, Hayashi PH, Protiva P, Kleiner DE, et al. Acute hepatitis E infection accounts for some cases of suspected drug-induced liver injury. Gastroenterology 2011;141:1665–1672.

[48] Avelar-Escobar G, Méndez-Navarro J, Ortiz-Olvera NX, Castellanos G, Ramos R, Gallardo-Cabrera VE, *et al*. Hepatotoxicity associated with dietary energy supplements: use and abuse by young athletes. Ann Hepatol 2012;11:564–569.

[49] Andrade RJ, Lucena MI, Fernández MC, Pelaez G, Pachkoria K, García-Ruiz E, *et al*. Spanish Group for the Study of Drug-induced Liver Disease. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. Gastroenterology 2005;129:512–521.

[50] Bessone F, Hernandez N, Dávalos M, Paraná R, Schinoni MI, Lizarzabal M, *et al*. Building a Spanish-Latin American network on drug induced liver injury; much to get from a joint collaborative initiative. Ann Hepatol 2012; 11:544–549.

[51] Chau TN, Cheung WI, Ngan T, Lin J, Lee KWS, Poon WT, *et al*. Causality assessment of herb-induced liver injury using multidisciplinary approach and the Roussel Uclaf Causality assessment Method (RUCAM). Clin Toxicol 2011; 49:34–39.

[52] García-Cortés M, Borraz Y, Lucena MI, Peláez G, Salmerón J, Diago M, *et al*. Liver injury induced by "natural remedies": an analysis of cases submitted to the Spanish Liver Toxicity Registry. Rev Esp Enferm Dig 2008;100:688–695. Article in Spanish.

[53] Teschke R, Schulze J. Suspected herbal hepatotoxicity: Requirements for appropriate causality assessment by the US Pharmacopeia. Drug Saf 2012; 35:1091–1097

[54] Stammschulte T, Gundert-Remy U. Spontaneous reports of primarily suspected herbal hepatotoxicity by Pelargonium sidoides: Was causality adequately ascertained? Regul Toxicol Pharmacol 2012;64:342.

[55] Teschke R, Frenzel C, Schulze J, Eickhoff A. Suspected herbal hepatotoxicity: The pharmacovigilance dilemma with disputed and obsolete evaluation methods. Regul Toxicol Pharmacol 2012;64:343–344.

[56] Manso G, López-Rivas L, Salgueiro ME, Duque JM, Jimeno FJ, Andrade RJ, *et al*. Continuous reporting of new cases in Spain supports the relationship between Herbalife products and liver injury. Pharmacoepidemiol Drug Saf 2011;20:1080–1087.

[57] Liss G, Lewis JH. Drug-induced liver injury: what was new in 2008? Expert Opin Drug Metab Toxicol 2009;5:843–860.

[58] Teschke R, Schmidt-Taenzer W, Wolff A. Spontaneous reports of assumed herbal hepatotoxicity by black cohosh: Is the liver unspecific Naranjo scale precise enough to ascertain causality? Pharmacoepidemiol Drug Saf 2011; 20:567–582.

[59] García-Cortés M, Lucena MI, Pachkoria K, Borraz Y, Hidalgo R, Andrade RJ. Evaluation of Naranjo Adverse Drug Reactions Probability Scale in causality assessment of drug-induced liver injury. Aliment Pharmacol Ther 2008;27: 780–789.

[60] Teschke R, Frenzel C, Schulze J, Eickhoff A. Spontaneous reports of primarily suspected herbal hepatotoxicity by *Pelargonium sidoides*: Was causality adequately ascertained? Regul Toxicol Pharmacol 2012;63:1–9.

[61] Teschke R, Frenzel C, Wolff A, Herzog J, Glass X, Schulze J, *et al*. Initially purported hepatotoxicity by *Pelargonium sidoides*: the dilemma of pharmacovigilance and proposals for improvements. Ann Hepatol 2012;11:500–512.

[62] Teschke R, Eickhoff A, Wolff A, Frenzel C, Schulze J. Herbal hepatotoxicity and WHO global introspection method. Ann Hepatol 2013;12:11–21.

[63] Teschke R, Glass X, Schulze J. Herbal hepatotoxicity by Greater Celandine (*Chelidonium majus*): Causality assessment of 22 spontaneous reports. Regul Toxicol Pharmacol 2011;61:282–291.

[64] Teschke R, Glass X, Schulze J, Eickhoff A. Suspected Greater Celandine hepatotoxicity: Liver specific causality evaluation of published case reports from Europe. Eur J Gastroenterol Hepatol 2012;24:270–280.

[65] Agarwal VK, McHutchison JG, Hoofnagle JH, Drug-Induced Liver Injury Network (DILIN). Important elements for the diagnosis of drug-induced liver injury. Clin Gastroenterol Hepatol 2010;8:463–470.